A U S H A N G

FREIE UNIVERSITÄT BERLIN

Fachbereich Mathematik und Informatik

Promotionsbüro, Arnimallee 14, 14195 Berlin

D I S P U T A T I O N

Donnerstag, 11. April 2019, 10 Uhr

Ort: FBR - Raum 1.1.16

(Fachbereich Physik, Arnimallee 14, 14195 Berlin)

Disputation über die Doktorarbeit von

Herrn Peter Hansen

Thema der Dissertation:

Robust algorithms for improved reproducible ChIP-seq and ChIP-nexus peak calling

Thema der Disputation:

Reproducible ChIP-seq peak calling

Die Arbeit wurde unter der Betreuung von **PD Dr. P. N. Robinson** durchgeführt.

Abstract: The high-throughput DNA sequencing application chromatin immunoprecipitation followed by high-throughput Sequencing (ChIP-seq) can be used to identify binding sites of target proteins in a genome wide fashion. The sequencing output typically consists of millions of short reads that first need to be mapped to a corresponding reference genome. The subsequent prediction of binding sites is referred to as peak calling. The Encyclopedia of DNA Elements (ENCODE) project consortium primarily used ChIP-seq to map hundreds of thousands of regulatory elements located in the non-coding part of the genome, whereby the reproducibility of ChIP-seq experiments performed by a number of different collaborators was evaluated using a framework called irreproducible discovery rate (IDR) procedure. In the first part of my presentation, I will present the IDR procedure. The first component of the IDR procedure is the correspondence curve, a graphical tool that can be used to visualize the transition from signal to noise for overlapping peaks derived from given pairs of replicates. The second component of this procedure is an inference procedure based on a copula mixture model that assigns peaks consistently called for pairs of replicates posterior probabilities of being irreproducible. These probabilities can then be used to control the rate of irreproducible peaks when selecting peaks at a given threshold, which is similar to the FDR. The second part of my presentation will summarize the key innovations arising from my thesis with a special focus on a saturation score for calling significant peaks in ChIP-seq data. Finally, I will briefly present a reproducibility comparison of different ChIP-seq peak callers that was performed within the IDR framework as well as some biologically relevant findings.

Die Disputation besteht aus dem o. g. Vortrag, danach der Vorstellung der Dissertation einschließlich jeweils anschließenden Aussprachen.

Interessierte werden hiermit herzlich eingeladen

Der Vorsitzende der Promotionskommission
Prof. Dr. K. Reinert